How Important are High Response Rates for College Surveys?

Kevin Fosnacht

Shimon Sarraf

Elijah Howe

Leah K. Peck

KEVIN FOSNACHT is an assistant research scientist at the Indiana University Center for Postsecondary Research and the National Survey of Student Engagement (NSSE). SHIMON SARRAF is the NSSE assistant director for survey operations and project services. ELIJAH HOWE and LEAH K. PECK are both Ph.D. candidates in the higher education program at Indiana University Bloomington. Earlier drafts of this paper were presented at the annual meetings of the American Educational Research Association, San Francisco, CA, April 2013 and the Association for Institutional Research, Long Beach, CA, May 2013. We would like to thank two anonymous reviewers for their helpful comments on earlier drafts. First authorship is shared by the first two alphabetically-listed contributors. Address queries to Shimon Sarraf at the Center for Postsecondary Research, Indiana University School of Education, 1900 East Tenth St., Suite 419, Bloomington, IN 47401; telephone: (812) 856–2169; email: ssarraf@indiana.edu.

Abstract

Higher education researchers have witnessed a long term decline in college student survey participation. However, survey methodologists have found that low response rates do not necessarily bias results. We investigated whether this proposition holds true for a college student survey by performing a level of effort analysis on 555 survey administrations that achieved a response rate of at least 50 percent. Using five quantitative measures, we found that population estimates, derived from a simulated sample with a low response rate, frequently are very similar to those based on actual high response rates.

Keywords: Response Rates, Non-response Bias, Student Engagement, National Survey of Student Engagement, Survey Research

How Important are High Response Rates for College Surveys?

Surveys play an important role in understanding the higher education landscape. About 60 percent of the published research in major higher education journals utilized survey data (Pike, 2007). Institutions also commonly use surveys to assess student outcomes and evaluate programs, instructors, and even cafeteria food. However, declining survey participation rates threaten this source of vital information and its perceived utility. Survey researchers across a number of social science disciplines in America and abroad have witnessed a gradual decrease in survey participation over time (Brick & Williams, 2013; National Research Council, 2013). Higher education researchers have not been immune from this trend; Dey (1997) long ago highlighted the steep decline in response rates in the American Council on Education and Cooperative Institutional Research Program (CIRP) senior follow-up surveys from 60 percent in the 1960s to 21 percent in 1991.

Survey researchers have long assumed that the best way to obtain unbiased estimates is to achieve a high response rate. For this reason, the literature on survey methods is rife with best practices and suggestions to improve survey response rates (e.g., American Association for Public Opinion Research, n.d.; Dillman, 2000; Heberlein & Baumgartner, 1978). These methods can be costly or require significant time or effort by survey researchers and may be unfeasible for postsecondary institutions due to the increasing fiscal pressures placed upon them. However, many survey researchers have begun to question the widely held assumption that low response rates provide biased results (Curtin, Presser, & Singer, 2000; Groves, 2006; Keeter, Miler, Kohut, Groves, & Presser, 2000; Massey & Tourangeau, 2013; Peytchev, 2013).

This study investigates this assumption with college student assessment data. It utilizes data from hundreds of samples of first-year and senior students with relatively high response rates using a common assessment instrument with a standardized administration protocol. It investigates how population estimates would have changed if researchers put forth less effort when collecting data and achieved lower response rates and respondent counts. Due to the prevalence of survey data in higher education research and assessment efforts, it is imperative to better understand the relationship between response rates and data quality.

Literature Review

Survey nonresponse bias—the extent to which survey nonresponse leads to inaccurate population estimates—has received extensive attention in the survey research literature (e.g., Curtin et al., 2000; Groves, 2006; Groves & Peytcheva, 2008; Rubin, 1976). Though variation exists with defining *nonresponse bias*, most view it as a function of the *response rate* and *nonresponse effect*, or how much responders and nonresponders differ on survey variables of interest (Keeter et al., 2000). In other words, low response rates may or may not lead to nonresponse bias because answers to survey items may not differ substantially between responders and nonresponders. The impact of nonresponse on an estimate depends upon the relationship between the outcome of interest and the decision to participate in the survey (Groves 2006). Consequently, if survey participation is not correlated with its content, the answers of responders and non-responders to a survey will not substantially differ. For these reasons, Massey and Tourangeau (2013) state that a high rate of nonresponse increases the *potential* for biased estimates, but does not necessarily bias an estimate. Peytchev (2013) goes further and argues that the use of response rate as the singular measure of survey representativeness is flawed, as “it is nonresponse bias that is feared, not nonresponse itself” (p. 89). This ambiguity with nonresponse bias was made explicitly clear when Keeter (2012) reported to the National Science Foundation that, “…there is no comprehensive theory of survey response that can generate reliable predictions about when nonresponse bias will occur" (p. 43). One can infer then that it is incumbent upon researchers to engage in nonresponse bias assessments since no one knows for sure whether bias will exist while using any given instrument with any population of interest.

Due to these insights, survey researchers have increasingly examined the impact of nonresponse on their survey estimates. Perneger, Chamot, and Bovier (2005) assessed nonresponse bias by comparing outcomes between early-, late-, and non-responders. They found a modest difference in their estimated outcomes (less than .1 standard deviations) when comparing population estimates based on samples with only early responders (30% response rate) and the full sample (70% response rate). The authors concluded that while nonresponse bias did exist, greater survey participation "has only minimal influence on the conclusions of the survey" (p. 380). Similarly, using data from the Index of Consumer Sentiment (ICS), Curtin and colleagues (2000) found no difference in their population estimates when comparing preliminary results based on response rates 5 to 50 percentage points lower than the final response rate. They created alternative estimates by excluding respondents that initially refused, required more than five recruitment calls, and required more than two recruitment calls. This analytical approach to assess population estimates under different response rate scenarios is generally referred to as a "level of effort" analysis (Olson, 2006), a term reflecting that a final response rate is somewhat artificial and dependent on when survey administrators stop contacting nonrespondents (or putting forth effort). Other health and psychology studies have come to similar conclusions based on results showing little variation under different response rate assumptions (Gerrits, van den Oord, & Voogt, 2001; Locker, 1993).

The results from these studies are not especially surprising given that other studies have found few differences between responders and nonresponders. Without a nonresponse effect, population estimates under different response rate scenarios should be highly correlated to estimates based on higher, final response rates. For instance, one study focusing on eating disorders determined that survey responses between first responders and those requiring several contacts did not differ (Mond, Rodgers, Hay, Owen, & Beumont, 2004). Additionally, a level of effort analysis of a telephone survey found minimal differences between the survey estimates on a wide variety of opinion, behavioral, and knowledge items from a study with a standard five-day field period and a more exhaustive study with an eight-week field period (Keeter et al., 2000). The significant differences observed were primarily for demographic items such as income and race.

In contrast, other researchers have found that increased efforts to collect survey data reduced nonresponse bias. One study, using household data from the German Panel Study, found that increased survey effort led to less nonresponse bias on a variety of individual characteristics (Kreuter, Muller, & Trappmann, 2010). Unlike the other studies above, they had administrative information for the entire sample so an *absolute* estimate of nonresponse bias could be calculated. This differs from *relative* estimates of nonresponse bias obtained from studies that do not utilize data with a 100 percent response rate. In addition, this study evaluated nonresponse bias by examining individual's background characteristics rather than less-tangible measures like an individual's perceptions or satisfaction. Another study came to the same conclusion when examining patient satisfaction data on ratings of physicians and found substantial differences in their estimated outcomes (Mazor, Clauser, Field, Yood, & Gurwitz, 2002). Comparing the final population estimate to one of three simulated estimates, they found almost a full standard deviation difference, suggesting the potential for substantial nonresponse bias.

Others have found that nonresponse had varying effects on population estimates by comparing survey data on school characteristics to the same characteristics gathered from a secondary data source (Kano, Franke, Afifi, & Bourque, 2008). The authors found significant differences between responders and nonresponders for two of the seven variables studied (population density and enrollment in English Learner programs). However, one of the variables, population density, was the only variable significantly correlated with survey response, thus demonstrating that biased estimates occur when response propensity is correlated with an outcome. This study also found that high-effort respondents did not significantly differ from low-effort respondents and nonrespondents using study variables.

Distinct, but related to non-response bias is the difference between early and late responders. Most research on this topic has focused on surveys of patients and found that healthier patients are more likely to respond early (Gadkari, McHorney, Pedari, & Gowda, 2011; Paganini-Hill, Hsu, Chao & Ross, 1993; Yessis & Rathert, 2006). However, others have found no substantial difference between early and late responders in surveys examining a variety of topics (Borg & Tuten, 2003; Keeter et al., 2000; Lahaut et al., 2003; Mond et al., 2004; Welch & Barlau, 2013). The most likely reason for the disparate findings between studies examining early and late responders is the relationship between the survey topic and the population being studied. For instance, sick patients simply may be physically unable to respond to a survey, thus biasing results. In contrast, a more general survey examining job satisfaction may not have an inherent feature that causes differential rates of response with the study’s target population.

A handful of higher education studies have focused on assessing survey nonresponse effect and bias. One study, based on about 600 first-year students enrolled in different classes assigned to different survey samples, did not find meaningful differences in students' perceptions of their academic environment when comparing estimates from administrations with response rates of 100 and 35 percent (Hutchison, Tollefson, & Wigington, 1987). Another series of studies conducted telephone interviews with randomly selected students who were asked to take the National Survey of Student Engagement (NSSE) multiple times, but failed to do so (Kuh, n.d.; Sarraf, 2005). These studies indicated that nonresponders responded differently to about half the tested survey items; however, they did not investigate the impact of nonresponse bias on institution-level population estimates. The authors cautioned that specific results indicating nonresponders to be more engaged may be the result of social desirability bias or telephone mode effects and not caused by true differences between responders and nonresponders. A third line of research examined the effectiveness of using survey weights to reduce nonresponse bias (Dey, 1997). It found that survey weights, derived from a regression predicting survey response, markedly improved population estimates and reduced nonresponse bias.

Several other higher education studies (Korkmaz & Gonyea, 2008; Porter & Umbach, 2006; Porter & Whitcomb, 2005; Sax, Gilmartin, & Bryant, 2003; Sax, Gilmartin, Lee, & Hagedorn, 2008) have focused on student and school characteristics associated with responding to surveys. However, these studies did not estimate how this might influence population estimates while taking into consideration response rates and nonresponse effect.

Theory

Survey nonresponse bias is a function of the nonresponse rate and the difference in means on an outcome between the respondents and nonrespondents. This relationship can be mathematically expressed as follows:

$$Bias_{NR} = Y_R - \bar{Y} = \pi_{NR} * (\bar{Y}_R - \bar{Y}_{NR})$$

Where, *Y* equals the survey measure, *NR* represents nonrespondent, *R* is respondent, and π_{NR} is the nonresponse rate. Unbiased estimates occur when the nonresponse rate or difference in means between responders and non-responders is zero. As researchers typically do not observe outcomes for nonresponders, they have traditionally emphasized reducing the nonresponse rate as much as possible to avoid obtaining unbiased estimates. Yet, unbiased estimates may also be obtained under conditions of high nonresponse when an outcome does not differ between respondents and nonrespondents.

Individuals will respond to a survey if they believe the benefits of participation will outweigh the costs. Leverage-saliency theory posits that when deciding to participate individuals assess a survey's features (e.g., topic, monetary incentive, organization) and their prominence in the request to participate (Groves, Singer, & Corning, 2000). Therefore, the effort exerted by a survey researcher plays a significant role in whether an individual participates in the survey, as incentives, customizing recruitment messages, and increasing the number of survey invitations generally improves response rates (Goyder 1982; Groves, Presser, & Dipko, 2004; Heberlein & Baumgartner, 1978). Thus, a survey's response rate is a product of the characteristics of potential respondents, the survey, and their interactions.

Research and theory on nonresponse generally overlooks the importance of surveyor effort. To demonstrate its importance, consider the effort expended by the Census Bureau when collecting data for the decennial census and by a grocery store who asks a customer to take a survey when paying for their items. In the former, the Census Bureau expends extraordinary effort when collecting data by administering multiple mailings, publicizing their efforts in the media, and making in-person visits to collect data from nonresponders. In contrast, the store will typically ask the customer to respond once and may enter the respondent into a contest with a low probability of winning a monetary reward. Both of these surveys could easily change their characteristics by exerting more or less effort, which could result in a different response rate. Therefore, an individual's classification as a respondent or nonrespondent is somewhat artificial, as their status may change due to different levels of effort exerted by the researcher.

Study Goals and Research Questions

This study investigated how survey population estimates vary under different response rate and respondent count assumptions from hundreds of college student survey administrations at a wide variety of North American colleges and universities. We hope these findings help initiate a robust discussion about survey data quality indicators and the role they play within the higher education research community.

With these goals in mind, the following questions guided this study:

1) Do simulated *low response rate* survey estimates about college student engagement provide reliable information based on comparisons to actual high response rate estimates?
2) Do simulated *low respondent count* estimates provide reliable information based on comparisons to full sample estimates?
3) Do these results vary by survey administration size?

Methods

Data

We examined these research questions using data from NSSE, one of the most widely used higher education assessment instruments. NSSE is annually administered to random or census samples of first-year and/or senior students using a standard protocol at approximately 500 to 750 bachelor's degree-granting institutions (National Survey of Student Engagement, 2012). For analytical purposes, NSSE treats each class within an institution as an independent survey administration, thereby creating between 1,000 and 1,500 discrete samples annually. This study's sample included data from NSSE administrations between 2010 and 2012 that achieved a response rate greater than 50 percent and contained at least 20 respondents. 555 survey administrations from 307 institutions met these requirements. All respondents were invited to participate by email and complete NSSE online. Nonresponders to the original invitation received up to four additional emails reminding them to complete NSSE. Table 1 provides the distribution of all administration response rates by the number of students invited to take the survey and aggregated 2010 Basic Carnegie Classification. Response rates varied between 50 and 100 percent with a median of 57 percent. Most administrations meeting our inclusion criteria had less than 250 students in their sample.

Insert Table 1 about here

Our analyses focused on five NSSE measures, commonly referred to as the Benchmarks: Level of Academic Challenge (LAC), Active and Collaborative Learning (ACL), Student-Faculty Interaction (SFI), Enriching Educational Experiences (EEE), and Supportive Campus

Environment (SCE). These measures are composites of multiple survey items placed on a 100-point scale. Previous research has shown that these measures produce dependable group means from samples with as few as 50 students (Pike 2012).

Analyses

Our data analysis was descriptive by nature. We first calculated means for each of the measures at simulated response rates of 5, 10, 15, 20, 25, 30, and 35 percent for each survey administration included in the sample. We simulated the means by averaging the score of the initial respondents up to the response rate of interest. For example, if 100 students were invited to take the survey, the first five respondents, as measured by the time of survey submission, would be included in the simulated mean for a 5 percent response rate. It should be noted that our data is *not* simulated or hypothetical; rather, we used observed data to simulate or re-estimate population means that would have been obtained under different response rate conditions, such as a shortened data collection period.

For each measure, we correlated the simulated means with the full sample mean. This approach is analogous to comparing the outcomes of the same survey administered with different levels of effort. The different levels of effort were hypothetical in this study, but could have been the product of factors such as a shorter field period, fewer reminder emails, generic invitations, or reduced student participation incentives. The correlations at these response rates were calculated for very small ($20 < N < 250$), small ($250 \leq N < 500$), medium ($500 \leq N < 1{,}000$), and large ($N \geq 1{,}000$) administration sizes separately. We used a conservative correlation of .90 for evaluating reliability. We repeated these analyses using respondent count as the level of effort indicator. The study examined survey estimates using simulated respondent counts of 10, 25, 50, 75, 100, 150, and 200 students. As with the response rate approach, we examined correlations between the full sample and simulated means across all institutions and by administration size.

After examining the correlation between the estimates, we examined the accuracy and magnitude of difference between the full and simulated sample estimates using both the response rate and respondent count approaches by calculating the absolute standardized mean difference (ASMD). This entailed calculating the difference in means between the full and simulated samples, dividing it by the standard deviation of the full sample, and taking its absolute value. The formula for this statistic is as follows:

$$ASMD = \left|\frac{\overline{Y_F} - \overline{Y_S}}{SD_F}\right|$$

After calculating the ASMDs, we aggregated all results from the five measures and then classified each into one of three categories according to the magnitude of the difference. The categories were trivial (< .09 SDs), small (.10 to .30 SDs), and medium/large (> .30 SDs). We chose these categories to align with a previous analysis of Benchmarks' effect sizes rather than Cohen's suggested groupings (Cohen, 1988; National Survey of Student Engagement, n.d.). To simplify the results, we included all five Benchmarks in these analyses, unlike the correlational analyses that examined each Benchmark separately.

Results

We initially investigated the correlations between the simulated survey estimates at different levels of effort and the full sample mean (see Table 2) for all administrations combined. At a simulated response rate of 5 percent, the correlations between the simulated estimate and the full sample estimate for the five measures ranged from .64 to .89. After increasing the simulated response rate to 10 percent, all of the measures had correlations of .80 or higher. At 20 percent, the correlations for four of the measures exceeded .90, and the exception, SCE, nearly met this threshold at .89. Consequently, the full sample estimates were very similar to the simulated means at a 20 percent response rate. The correlations continued to rise along with the simulated response rate and approached 1.00 at a simulated rate of 35 percent.

Insert Table 2 about here

Next, we examined these correlations by administration size. Stronger correlations were observed between the simulated and full sample means as administration size increased. For very small administrations with less than 250 students, the correlation between the simulated means at a five percent response rate and the full sample means ranged from .58 and .84. In contrast, we observed correlations between .93 and .99 for the same measures among large administrations with at least 1,000 students. As with the overall results, the correlations between the simulated means and the full sample means rose along with the simulated response rate. The correlations for the very small administrations were greater than .90 for all measures using simulated means based on a 25 percent response rate. This bar was passed at a 10 percent simulated response rate for the large and medium administration sizes.

After examining the results by response rate, we replicated the above analyses by respondent counts (see Table 3). Using all administrations, the correlations between a mean derived from the first 10 respondents and the full sample ranged between .68 and .92 for the five measures. Correlations rose to between .86 and .97 using 25 respondents and exceeded .90 with 50 respondents. In contrast to the results by response rate, respondent count correlations did not vary substantially by administration size. For the four administration sizes, the correlations between a respondent count of 25 and the full sample mean ranged between .86 to .93, .85 to .94,

.95 to .97, and .80 to .88 for ACL, SFI, EEE, and SCE, respectively. The correlations between these measures were slightly less consistent for LAC, ranging from .74 to .90. Seventeen of the twenty correlations between the full sample mean and the means derived from the first 50 respondents exceeded .90. The three exceptions surpassed this threshold after raising the respondent count to 75 students.

Insert Table 3 about here

Next, we investigated the magnitude of the mean difference between the full and simulated samples. Table 4 contains the results when using different response rate assumptions to calculate the absolute standardized mean differences. When looking at all administrations, about one in ten of the mean differences exhibited a substantial difference (defined as an ASMD greater than .3 SDs) when the simulated response rate was 15%. This figured declined to one in twenty when the response rate was increased to 20%. However, like the correlations, the results differed by the size of the administration. Roughly half of the ASMDs for the very small administrations were greater than .3 SDs using a simulated response rate of 5%; however, only 1% of the large administrations exceeded this threshold at the same simulated response rate.

Insert Table 4 about here

When examining the magnitude of the mean differences using respondent counts to simulate different levels of surveyor effort, about one in ten means were substantially different when the respondent count was 25 students (see Table 5). Only two and one percent of the standardized mean differences were substantial when the respondent counts were increased to 50 and 75 students, respectively. The distributions using respondent counts to simulate different surveyor effort levels were roughly similar when examining the results by administration size.

Insert Table 5 about here

Discussion & Limitations

Response rates have been declining for both higher education and social science surveys generally for many years (Dey, 1997; National Research Council, 2013). This trend threatens the validity of higher education survey research and postsecondary accountability initiatives like the Voluntary System of Accountability. However, multiple prominent survey researchers have questioned the widely assumed relationship between low response rates and data quality (Curtin et al., 2000; Groves, 2006; Keeter et al., 2000; Massey & Tourangeau, 2013; Peytchev, 2013).

In this study, we examined the relationship between low response rates and the reliability of five measures widely used in higher education assessment by analyzing data from over 500 NSSE survey administrations with response rates of at least 50 percent. With few exceptions, we found estimates for several measures of college student engagement to be reliable under low response rate conditions (5% to 10%), provided the sampling frame included at least 500 students. For smaller administrations, the response rate required for an estimate to be reliable was higher, but we found estimates to be increasingly reliable after receiving responses from 50 to 75 students. Additionally, we examined the magnitude of the difference between simulated and full sample estimates. The results from these analyses comport with our reliability estimates in that few major differences were observed in the medium or large administrations at low simulated response rates or for smaller administrations when the respondent counts were 50 or 75 students. These findings support the work of Hutchison and colleagues (1987) that shows similar survey estimates of college student outcomes can be achieved based on a relatively low response rate administration, and also comport with previous NSSE non-responder studies (Kuh, n.d.; Sarraf, 2005), as well as Pike's (2012) findings that NSSE benchmark scores based on 50 respondents provide dependable group means.

These results suggest that institutions and researchers examining college student behavior may not need to exert great effort maximizing response rates. Rather, the level of effort exerted by an institution can be contingent upon the size of the student population being examined and the anticipated uses for the survey data. The results indicate that surveys with small sampling frames need a relatively high response rate (20 to 25 percent) to be fairly confident in their survey estimates. In contrast, surveys with large sampling frames can obtain reliable estimates with lower response rates. Regardless of administration size, a researcher's level of effort might be reduced, freeing time and monetary resources that could be better spent improving the survey instrument, analyzing the data or on other important projects. These considerations should be balanced with anticipated data usage. For example, more surveyor effort should be placed on achieving a high degree of accuracy when survey results will be used for accountability purposes (i.e., performance-based funding, teaching evaluations used for promotion and pay increases), particularly when examining subgroup results with fewer respondents, than for a survey seeking to evaluate a speaker or student food preferences. While we are hesitant to suggest hard rules for needed survey reliability, it is clear that a high stakes survey should be designed to have a reliability closer to 1.00 than a lower stakes survey. We echo Nunnally's (1978) commentary that a reliability of .70 may be sufficient for early or low stakes research purposes, but a reliability of .90 may not be sufficient for an applied setting with important consequences

The findings also suggest that researchers should pay more attention to other sources of potential error, besides nonresponse, when evaluating data quality. We share Peytchev’s (2013) concern that the overwhelming attention received by the response rate might distract from attending to other important types of survey error, such as measurement and sampling error. More emphasis should also be placed on investigating other data quality measures such as response differentiation, survey duration, and item nonresponse.

One important issue to review is whether the level of effort put forth by survey administrators should be guided by response rates or respondent counts. These results suggest that if you had to choose one, focusing on respondent counts would be wise, regardless of sample size. As stated previously, 50 to 75 respondents provided reliable estimates for the five measures, whereas the response rate needed to achieve reliable estimates varied considerably by administration size. Focusing solely on response rates may lead to confusion for large survey administrators like NSSE or CIRP due to the varied response rates required across different administration sizes. However, response rates play a prominent role in data quality determinations by many constituents; therefore, response rates cannot be dismissed as irrelevant. As many know well, characterizing any individual survey administration as suffering from a low response rate will influence how results are received, regardless of how many individuals respond to a survey. Some governmental organizations, in fact, require that a survey achieve a certain response rate (Office of Management and Budget, 2006). If it does not, researchers must conduct additional analyses to ensure the representativeness of the data. Given the importance that many place on response rates, it is probably wise for most survey administrators and higher education researchers to continue to utilize proven methods that increase participation to ensure their findings are not dismissed by their intended audiences.

Porter, Whitcomb, & Weitzer (2004) presciently forecasted over ten years ago that decreasing costs associated with designing and administering online surveys would make survey fatigue more prevalent. As many colleges and universities struggle with this phenomenon today, they may want to consider randomly sampling smaller groups of students when administering surveys. Hypothetically, if the aim is to collect 50 respondents for a reliable estimate, and your population is 1,000, a reasonable approach would be to randomly sample 200 students, assuming a 25 percent response rate. The remaining 800 unsampled students could be used for other research or assessment projects, thus reducing survey fatigue and potentially increasing response rates for all surveys being administered on campus. This approach would require researchers to be more strategic with planning surveys for their campus, as well as requiring them to make accurate projections for ensuring a minimum respondent count. Large survey administrators, such as NSSE and CIRP, might also consider calculating for institutions an optimal sample size to yield a minimum number of respondents. Though coordinated sampling could ameliorate survey fatigue, strong, centralized survey coordination will be required as well given the plethora of offices and individuals surveying students on many campuses. Support for minimizing survey fatigue by delivering fewer survey requests can also be found in several other studies. Porter and colleagues (2004) found that surveys that quickly follow a previous survey show lower response rates, but larger time gaps between individual administrations exhibited weaker effects. In addition, others have found that increased online surveying to elicit teacher evaluations from college students corresponded with eventual decreased participation (Adams & Umbach, 2012). With these findings in mind, it is our opinion that the amount of effort expended on each survey be balanced with how strongly each survey relates to a college or university's primary mission: educating students.

Despite the strong rationale for limiting the size of a survey administration through random sampling, researchers should recognize that this approach presents some risks and carefully assess their need to maximize response. Significantly fewer respondents will lead to less precise population estimates and a greater probability of making a Type-II error when conducting statistical comparisons. Fewer respondents also means less data and power to investigate various student sub-groups on campus (e.g., academic major, ethnicity) and less confidence in these estimates. Before adopting a random rather than census sampling approach, researchers should anticipate all possible impacts this might have on analyzing student sub-groups or needed statistical analyses.

A few study limitations should be noted before drawing any final conclusion. First, the study examined relative, not absolute, nonresponse bias. In other words, despite using relatively high response rate administrations in this study, knowing the true population statistic for all administrations could influence our results in some unanticipated way. Second, the survey administrations meeting our 50 percent response rate criteria may be unique in a way that strongly influences our findings. For instance, the mean difference between early and late responders among schools with less than 50 percent response rates may be greater than the difference between these two groups at institutions within our study, thus resulting in lower reliability between simulated results and actual results.

Future investigations should help to shed light on identifying administrations that do not demonstrate reliable survey estimates with few respondents or low response rates. Nonresponders (or late responders) at some institutions may actually be very different than responders (or early responders), in which case exerting as much effort at boosting overall response rates and respondent counts would be warranted. In addition, determining the generalizability of this study's findings across other college assessment instruments is important. We believe our results are a function of *general* student interest in NSSE's content and that other survey topics may not yield similar results. Consistent with leverage-saliency theory, research conducted by Groves, Presser, & Dipko (2004) suggest the likelihood of responding to various surveys is the result of an interaction between survey topic and an individual's background. For example, a survey about dormitory food is more likely to be completed by on-campus than off-campus residents. This provides an explanation for why students of varying engagement levels appear to have proportional likelihoods to complete the survey at various points in time during an administration (at most institutions), thus minimizing the risk of unreliable population estimates at low response rates and respondent counts. Most likely the vast majority of students at any particular campus find NSSE's content equally relevant to them. Additionally, further research in higher education should focus on alternative measures of survey representativeness to the response rate. Some alternative measures suggested by a number of prominent survey researchers include: goodness of fit statistics, R-indices, the fraction of missing information, and the variance of nonresponse weights and rates by groups (Groves et al., 2008).

Conclusion

Researchers focusing on college students and that want to increase their response rates to an arbitrary number to satisfy preconceived notions of a "good" response rate should question whether their extra effort is warranted. This study did not find that a 5% response rate or even a 75% response rate provides unbiased population estimates under all circumstances, but rather that additional effort to move response rates marginally higher will frequently only shift survey results in trivial ways after one collects a minimum number of responses. Once researchers consider these results, they may spend less time worrying about achieving a high response rate and more time evaluating and using the data they collect.

References

Adams, M., & Umbach, P. D. (2012). Nonresponse and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education*, *53*(5), 576-591.

American Association for Public Opinion Research. (n.d.) *Best practices*. Retrieved from http://www.aapor.org/Best_Practices1.htm.

Borg, I., & Tuten, T. L. (2003). Early versus later respondents in intranet-based, organizational surveys. *Journal of Behavioral and Applied Management*, *4*(2), 134-147.

Brick, J. M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American academy of political and social science*, *645*, 36-59.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, *64*(4), 413-428.

Dey, E. L. (1997). Working with low survey response rates: The efficacy of weighting adjustments. *Research in Higher Education*, *38*(2), 215-227.

Dillman, D. A. (2000). *Mail and Internet surveys: The tailored design method.* (2nd ed.) New York: J. Wiley.

Gadkari, A., McHorney, C. Pedari, A., & Gowda, N. (2011). Non-response bias in a survey assessing patients' medication and health beliefs. *Value in Health*, *14*(3), A149.

Gerrits, M. H., van den Oord, E., & Voogt, R. (2001). An evaluation of nonresponse bias in peer, self, and teacher ratings of children's psychosocial adjustment. *Journal of Child Psychology and Psychiatry*, *42*(5), 593-602.

Goyder, J. C. (1982). Further evidence on factors affecting response rates to mailed questionnaires. *American Sociological Review*, *47*(4), 550-553.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys, *Public Opinion Quarterly*, *70*(5), 646-675.

Groves, R. M., Brick, J. M., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Kreuter, F., ..., & Wagner, J. (2008). Issues facing the field: Alternative practical measures of representativeness of survey respondent pools. *Survey Practice*, *1*(3).

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–89.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, *68*(1), 2–31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation: Description and an Illustration. *Public Opinion Quarterly*. *64*(3), 299-308.

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, *43*(4), 447-462.

Hutchison, J., Tollefson, N., & Wigington, H. (1987). Response bias in college freshman’s responses to mail surveys. *Research in Higher Education*, *26*(1), 99–106.

Kano, M., Franke, T., Afifi, A. A., & Bourque, L. B. (2008). Adequacy of reporting results of school surveys and nonresponse effects: A review of the literature and a case study. *Educational Researcher*, *37*(8), 480-490.

Keeter, S. (2012). The impact of survey nonresponse on survey accuracy. In J. A. Krosnick, S. Presser, K. Husbands Fealing, & S. Ruggles (Eds.), The future of survey research: Challenges and opportunities (pp. 42-43). Washington, DC: The National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, *64*(2), 125-148.

Korkmaz, A., & Gonyea, R. M. (2008). *The effect of precollege engagement on the likelihood of response to the National Survey of Student Engagement*. Paper presented at the annual meeting of the Association for Institutional Research, Seattle, WA.

Kreuter, F., Muller, G., & Trappmann, M. (2010). Nonresponse and measurement error in employment research. *Public Opinion Quarterly*, *74*(5), 880-906.

Kuh, G. D. (n.d.). *The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties*. Retrieved from http://nsse.iub.edu/pdf/conceptual_framework_2003.pdf

Lahaut, V. M., Jansen, H. A., van de Mheen, D., Garretsen, H. F., Verdurmen, J. E., & Van Dijk, A. (2003). Estimating non-response bias in a survey on alcohol consumption: Comparison of response waves. *Alcohol and Alcoholism*, *38*(2), 128-134.

Locker, D. (1993). Effects of non-response on estimates derived from an oral health survey of older adults. *Community Dentistry and Oral Epidemiology*, *21*(2), 108-113.

Massey, D. S., & Tourangeau, R. (2013). Where do we go from here? Nonresponse and social measurement. *The ANNALS of the American Academy of Political and Social Science, 645*(1), 222-236.

Mazor, K. M., Clauser, B. E., Field, T., Yood, R. A., & Gurwitz, J. H. (2002). A demonstration of the impact of response bias on patient satisfaction surveys. *Health Services Research*, *37*(5), 1403-1417.

Mond, J. M., Rodgers, B., Hay, P. B., Owen, C., & Beumont, P. J. V. (2004). Nonresponse bias in a general population survey of eating-disordered behavior. *International Journal of Eating Disorders, 36*(1), 89-98.

National Research Council. (2013). *Nonresponse in Social Science Surveys: A Research Agenda*. Washington, DC: The National Academies Press.

National Survey of Student Engagement. (n.d.). *Contextualizing NSSE effect sizes: Empirical analysis and interpretation of benchmark comparisons*. Retrieved from http://nsse.indiana.edu/pdf/effect_size_guide.pdf

National Survey of Student Engagement. (2012). *Administration protocol and procedures*. Retriveved from http://nsse.indiana.edu/html/administration_protocol.cfm.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Office of Management and Budget. (2006). *Standards and guidelines for statistical surveys*. Retrieved from https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias and total bias. *Public Opinion Quarterly, 70*(5), 737-758.

Paganini-Hill, A., Hsu, G., Chao, A., & Ross, R. K. (1993). Comparison of early and late respondents to a postal health survey questionnaire. *Epidemiology*, *4*(4), 375-379.

Perneger, T. V., Chamot, E., & Bovier, P. A. (2005). Nonresponse bias in a survey of patient perceptions of hospital care. *Medical Care*, *43*(4), 374-380.

Peytchev, A. (2013). Consequences of survey nonresponse. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 88-111.

Pike, G. R. (2007). Adjusting for nonresponse in surveys. In John C. Smart (Ed.), *Higher education: Handbook of theory and research, vol. 22* (pp. 411-450). The Netherlands: Springer.

Pike, G. R. (2012). NSSE benchmarks and institutional outcomes: A note on the importance of considering the intended uses of a measure in validity studies. *Research in Higher Education*, *54*(2), 149-170.

Porter, S. R., & Umbach, P. D. (2006). Student survey response rates across institutions: Why do they vary? *Research in Higher Education*, *47*(2), 229-247.

Porter, S. R., & Whitcomb, M. E. (2005). Nonresponse in student surveys: The role of demographics, engagement and personality. *Research in Higher Education*, *46*(2), 127-152.

Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, *121*, 63-73.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3) 581-592.

Sarraf, S. A. (2005). *An evaluation of nonresponse effect in the National Survey of Student Engagement*. Bloomington, IN: Center for Postsecondary Research, Indiana University.

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, *44*(4), 409-432.

Sax, L. J., Gilmartin, S. K., Lee, J., & Hagedorn, L. S. (2008). Using web surveys to reach community college students: An analysis of response rates and response bias. *Community College Journal of Research and Practice*, *32*(9), 712-729.

Welch, W. W., & Barlau, A. B. (2013). Addressing survey nonresponse issues: Implications for ATE principal investigators, evaluators, and researchers. Retrieved from http://www.colorado.edu/ibs/decaproject/pubs/Survey%20nonresponse%20issues%20Implications%20for%20ATE%20PIs%20researchers%20%20evaluators.pdf

Yessis, J., & Rathert, C. (2006). Initial versus prompted responders to patient satisfaction surveys: Implication for interpretation of patient feedback. *Journal of Applied Management and Entrepreneurship*, *11*(4), 49.

Table 1
Response rate distribution of administrations included in the study by administration size[1] and Carnegie Classification

		Percentile						
	N	Min.	10	25	50	75	90	Max.
Administration Size								
Very Small	293	50	51	53	59	67	74	100
Small	168	50	51	53	55	60	68	76
Medium	74	50	51	52	55	60	64	76
Large	20	50	50	50	52	55	69	72
Carnegie Classification (aggregated)								
Baccalaureate	335	50	51	54	58	64	71	98
Master's	117	50	50	52	54	59	61	100
Doctoral	13	50	50	51	55	60	65	65
Other/Not Classified	90	50	51	52	59	68	78	94
Total	**555**	**50**	**51**	**53**	**57**	**63**	**71**	**100**

[1] Administration size is the number of students asked to take NSSE.

Note: Very small = Less than 250 students sampled; Small = 250 through 499 students sampled; Medium = 500 through 999 students sampled; Large = 1,000 or more students sampled

Table 2
Correlations between simulated response rate and full sample means by survey measure and administration size

	Simulated Response Rate						
	5	10	15	20	25	30	35
Level of Academic Challenge							
All administrations	.64	.80	.86	.91	.93	.95	.97
Very small	.61	.76	.83	.90	.92	.94	.96
Small	.69	.87	.91	.95	.95	.96	.98
Medium	.78	.91	.94	.95	.97	.98	.99
Large	.94	.98	.98	.99	.99	.99	.99
Active and Collaborative Learning							
All administrations	.76	.88	.93	.95	.96	.97	.98
Very small	.69	.82	.89	.93	.95	.96	.97
Small	.79	.90	.94	.95	.97	.98	.98
Medium	.93	.97	.98	.99	.99	.99	1.00
Large	.97	.99	.99	.99	1.00	1.00	1.00
Student-Faculty Interaction							
All administrations	.75	.87	.92	.95	.96	.97	.98
Very small	.68	.81	.89	.92	.95	.96	.97
Small	.82	.91	.95	.97	.97	.98	.99
Medium	.89	.96	.98	.99	.99	.99	.99
Large	.93	.98	.98	.99	.99	.99	.99
Enriching Educational Experiences							
All administrations	.89	.95	.97	.98	.98	.99	.99
Very small	.84	.92	.96	.97	.98	.98	.99
Small	.94	.97	.98	.99	.99	1.00	1.00
Medium	.97	.99	.99	1.00	1.00	1.00	1.00
Large	.99	.99	.99	.99	1.00	1.00	1.00
Supportive Campus Environment							
All administrations	.66	.80	.86	.89	.92	.95	.96
Very small	.58	.74	.81	.86	.90	.93	.95
Small	.79	.89	.94	.95	.95	.97	.98
Medium	.84	.90	.93	.94	.96	.97	.98
Large	.95	.97	.98	.98	.99	.99	1.00

Note: Very small = Less than 250 students sampled; Small = 250 through 499 students sampled; Medium = 500 through 999 students sampled; Large = 1,000 or more students sampled

Table 3
Correlation between simulated respondent count and full sample means by survey measure and administration size

	Simulated Respondent Count						
	10	25	50	75	100	150	200
Level of Academic Challenge							
All administrations	.68	.87	.94	.96	.97	.98	.99
N	*555*	*551*	*494*	*430*	*362*	*245*	*178*
Very small	.74	.90	.96	.98	.99	.99	---
N	*293*	*289*	*232*	*168*	*100*	*8*	*0*
Small	.58	.83	.92	.95	.97	.99	.99
N	*168*	*168*	*168*	*168*	*168*	*143*	*83*
Medium	.57	.74	.88	.91	.94	.97	.98
N	*74*	*74*	*74*	*74*	*74*	*74*	*74*
Large	.55	.76	.92	.94	.97	.97	.98
N	*20*	*20*	*20*	*20*	*20*	*20*	*20*
Active and Collaborative Learning							
All administrations	.81	.92	.96	.97	.98	.99	.99
N	*555*	*551*	*494*	*430*	*362*	*245*	*178*
Very small	.82	.93	.97	.98	.99	1.00	---
N	*293*	*289*	*232*	*168*	*100*	*8*	*0*
Small	.69	.86	.94	.96	.97	.99	1.00
N	*168*	*168*	*168*	*168*	*168*	*143*	*83*
Medium	.85	.92	.96	.98	.98	.99	1.00
N	*74*	*74*	*74*	*74*	*74*	*74*	*74*
Large	.84	.91	.93	.96	.97	.97	.98
N	*20*	*20*	*20*	*20*	*20*	*20*	*20*
Student-Faculty Interaction							
All administrations	.79	.92	.96	.98	.98	.99	.99
N	*555*	*551*	*494*	*430*	*362*	*245*	*178*
Very small	.82	.94	.97	.98	.99	1.00	---
N	*293*	*289*	*232*	*168*	*100*	*8*	*0*
Small	.70	.89	.95	.97	.98	.99	1.00
N	*168*	*168*	*168*	*168*	*168*	*143*	*83*
Medium	.77	.85	.95	.96	.97	.99	.99
N	*74*	*74*	*74*	*74*	*74*	*74*	*74*
Large	.78	.87	.90	.96	.95	.96	.97
N	*20*	*20*	*20*	*20*	*20*	*20*	*20*
Enriching Educational Experiences							
All administrations	.92	.97	.98	.99	1.00	1.00	1.00
N	*555*	*551*	*496*	*434*	*363*	*247*	*176*
Very small	.92	.97	.99	1.00	1.00	1.00	---
N	*293*	*289*	*234*	*172*	*101*	*10*	*0*
Small	.89	.96	.98	.99	.99	1.00	1.00
N	*168*	*168*	*168*	*168*	*168*	*143*	*82*
Medium	.94	.97	.98	.99	.99	1.00	1.00
N	*74*	*74*	*74*	*74*	*74*	*74*	*74*
Large	.84	.95	.97	.99	.99	.99	.99
N	*20*	*20*	*20*	*20*	*20*	*20*	*20*

Table 3 (continued)

	Simulated Respondent Count						
	10	25	50	75	100	150	200
Supportive Campus Environment							
All administrations	.70	.86	.93	.95	.96	.97	.98
N	*555*	*551*	*494*	*430*	*362*	*245*	*178*
Very small	.70	.88	.95	.98	.99	.96	---
N	*293*	*289*	*232*	*168*	*100*	*8*	*0*
Small	.71	.84	.93	.95	.97	.99	.99
N	*168*	*168*	*168*	*168*	*168*	*143*	*83*
Medium	.62	.80	.89	.91	.93	.96	.97
N	*74*	*74*	*74*	*74*	*74*	*74*	*74*
Large	.75	.86	.84	.91	.92	.95	.96
N	*20*	*20*	*20*	*20*	*20*	*20*	*20*

Note: Very small = Less than 250 students sampled; Small = 250 through 499 students sampled; Medium = 500 through 999 students sampled; Large = 1,000 or more students sampled

Table 4.
Distribution of absolute standardized mean differences between full sample and simulated response rate samples by administration size

	Simulated Response Rate						
Absolute Standardized Mean Differences	5	10	15	20	25	30	35
All administrations							
Trivial (<.1 SDs)	27	40	49	59	65	72	79
Small (.1-.3 SDs)	41	43	42	36	32	27	21
Medium/Large (> .3 SDs)	33	17	9	5	3	2	1
Very small							
Trivial (<.1 SDs)	19	30	37	47	52	59	67
Small (.1-.3 SDs)	36	43	47	44	43	38	32
Medium/Large (> .3 SDs)	45	26	16	9	6	3	1
Small							
Trivial (<.1 SDs)	31	47	59	69	76	82	89
Small (.1-.3 SDs)	45	45	39	30	23	18	11
Medium/Large (> .3 SDs)	24	8	2	1	1	0	0
Medium							
Trivial (<.1 SDs)	40	54	65	77	86	91	97
Small (.1-.3 SDs)	48	44	35	24	14	9	3
Medium/Large (> .3 SDs)	12	2	1	0	0	0	0
Large							
Trivial (<.1 SDs)	60	81	82	91	92	93	95
Small (.1-.3 SDs)	39	19	18	9	8	7	5
Medium/Large (> .3 SDs)	1	0	0	0	0	0	0

Notes: Values are percentages. Includes all five NSSE Benchmarks. Very small = Less than 250 students sampled; Small = 250 through 499 students sampled; Medium = 500 through 999 students sampled; Large = 1,000 or more students sampled

Table 5.
Distribution of absolute standardized mean differences between full sample and simulated count samples by administration size

	Simulated Respondent Count						
Absolute Standardized Mean Differences	10	25	50	75	100	150	200
All administrations							
Trivial (<.1 SDs)	25	45	62	74	81	89	94
Small (.1-.3 SDs)	41	44	36	26	19	11	6
Medium/Large (> .3 SDs)	34	11	2	1	0	0	0
Very small							
Trivial (<.1 SDs)	26	49	71	85	96	93	---
Small (.1-.3 SDs)	42	44	28	15	4	7	---
Medium/Large (> .3 SDs)	32	7	1	0	0	0	---
Small							
Trivial (<.1 SDs)	25	42	58	71	82	96	99
Small (.1-.3 SDs)	37	44	40	29	18	5	1
Medium/Large (> .3 SDs)	38	14	3	0	0	0	0
Medium							
Trivial (<.1 SDs)	22	36	49	60	64	83	93
Small (.1-.3 SDs)	45	47	48	39	35	17	7
Medium/Large (> .3 SDs)	32	17	3	1	0	0	0
Large							
Trivial (<.1 SDs)	22	39	43	53	64	66	76
Small (.1-.3 SDs)	48	48	50	45	36	34	24
Medium/Large (> .3 SDs)	30	13	7	2	0	0	0

Notes: Values are percentages. Includes all five NSSE Benchmarks. Very small = Less than 250 students sampled; Small = 250 through 499 students sampled; Medium = 500 through 999 students sampled; Large = 1,000 or more students sampled